# CREATING MULTI-PAGE DOCUMENTS USING TIFF FILES

## Reference to Provisional Application

The present application claims priority from US Provisional Application
5  Serial No. 60/168,293, filed December 1, 1999.

## Field of the Invention

The present invention is directed to a method for representing a multi-
page document using a hierarchically-organized set of TIFF files.

10

## Background of the Invention

"TIFF-FX" is a proposed standard for the rendering and retention of image
data.  It is useful for transmission of facsimile-format documents over the
Internet, and encompasses other standards such as JPEG, JBIG, and color fax
15  standards.  One aspect of TIFF-FX is that there is a special problem with
rendering multi-page documents, and/or page images having multiple
components (such as combinations of text, contone images, and line art) in a
coherent format.

In TIFF-FX, different types of image components (text, line art, contone)
20  can be compressed in various ways, such as JBIG, JPEG, or fax formats.  The
different compression arrangements or schemes are called "profiles."  Examples
of profiles are:

S= b/w, simple compression algorithm

F= b/w, richer compression algorithm

25  J= b/w, JBIG compression

C= color JPEG compression

L= color JBIG compression

M= MRC = "mixed raster content" = in each page, different components are compressed in different ways. Different components of a page image are organized as "mask," "upper," and "lower," which are ultimately combined to create a single, multi-component page image. Typically, the "mask" is text, compressed in binary, JBIG, or the fax compressions Modified Huffman, Modified Read, or Modified Modified Read. The "lower" portion is typically contone images compressed in JPEG. The "upper" portion is typically line art compressed in GZIP.

The present invention is directed to a system for organizing image data in a heterogeneous form, such as including both color and monochrome images, or images compressed according to different schemes, so that a TIFF-FX writer can automatically organize the data to create a single multi-page document.

## Description of the Prior Art

U.S. Patent **5,706,457** discloses a system for acquiring and archiving images derived from multiple sources. An operator of the system can perform only a predetermined set of functions corresponding to graphical icons. Each of the icons launches a set of macro functions that format the image data into a predetermined format.

U.S. Patent **5,724,579** discloses a system for producing "subordinate images" extracted from our original image data. The subordinate image data can be images directed to a portion of the original data, or subset of the original data making a thumbnail of the original data. A first subordinate image is extracted from original image data, and a second subordinate image is in turn extracted from the first subordinate image data. The main image data and the first and second subordinate image data are stored in the same file.

U.S. Patent **6,052,198** discloses a system for organizing files associated with a single job ticket, such as in a digital printing context. The job ticket includes information on print files included in a print job, print file location information indicating a location of print files in a storage device, and information indicating a location of a rasterized of version of a print file in the storage device.

When the job ticket is submitted to a printing apparatus, a rasterized version of the data is submitted instead of the original print file if the rasterized version was modified after the print file was modified.

## Summary of the Invention

According to one aspect of the present invention, there is provided a method of organizing image data to create a multi-page document, comprising the steps of naming each of a set of files, each file representing either a page image or an image component of a page image, according to a naming convention, organizing the files into a hierarchical directory structure, and applying a writer application which recognizes the files by the naming convention to create a single, multi-page document.

## Brief Description of the Drawing

Figure 1 describes an input file organization according to an embodiment of the present invention.

## Detailed Description of the Invention

According to the present invention, a naming convention and directory structure is used to identify individual page images and/or page image components within a multi-page document, so that a multi-page, multi-component "source file" can be created. The basic approach to converting many single page TIFF files into a single TIFF-FX file is to: (1) organize the original TIFF files into a specified architecture (by a combination of naming convention and directory structure); and (2) execute a known TIFF-FX writer application to convert the set of TIFF files into the TIFF-FX file.

The input data to the writer must be in a particular hierarchy on a disk to be properly handled. Figure 1 describes the input file organization. It can be seen in the Figure that a TIFF-FX writer can recognize a single page input file, a directory of input files (for a multi-page document of simple pages), or as a directory of directories (for a multi-page document wherein some or all pages

have multiple components, as described above). Quality Logic (formerly Genoa Systems) currently sells a product "TIFF-FXpert Test System" used to evaluate TIFF-FX files: this product can be used as a writer within the context of the present invention.

5          According to one embodiment of the present invention, there are three modes to the TIFF-FX writer.  If the "source argument" (the name of the file desired to be considered a single document) is a simple file name, then a single page TIFF-FX file will be generated.  In such a case, any profile may be requested except the MRC profile, M.  If the source argument is a directory of

10  files, then a multi-page TIFF-FX document is generated.  According to the convention of one embodiment, each file in the directory must have a file name "PageN" where N is a page number starting with 1.  Source files not obeying this convention are ignored.  Once again, in this case any profile may be requested except the MRC profile, M.

15          To support the MRC profile, M, the source argument, may represent a directory of page directories.  Each file must contain at least three files, which, in one convention, are named Mask, Lower, and Upper, corresponding to the roles described above in MRC profile layers.  To test profile M, input data follow this format. All other files will be ignored.

20          According to one alternate embodiment, there may also be included, in the hierarchy, an "info" or "directive" file, which contains data relating to at least some of the other files within the same directory.  This "info" file could include instructions that, for instance, the text in the mask within the same directory should be compressed in a specific way, such as in G3 format, or the contone

25  data must be compressed in JPEG; also, the info file can specify a particular quality level for the compression algorithm.

          In a preferred embodiment, all source files should conform to TIFF6 (baseline + standard extensions) specifications.

          With all source images in the format described above, the TIFF-FX writer

30  can proceed to read original data in various formats and emit the hierarchically-organized TIFF-FX files.

Although a TIFF-FX implementation is shown here, the basic principle can be applied to the creation of other multi-page document formats.

The present invention simplifies the testing and debugging of TIFF-FX images. TIFF-FX files can potentially represent many pages of image data, each page being quite complex (i.e., profile M). Real applications may require significant additional processing (e.g., segmentation of an image into Foreground, Background, and Mask layers). This representation allows separation of the development of segmentation algorithms from the development of the TIFF-FX writers/readers, and defines a common means by which developers can interchange test data. The present invention can be used to convert existing repositories of document data into TIFF-FX files. Scripts can be constructed that would take existing repositories and convert them into the appropriate hierarchy, then a TIFF-FX writer would generate the TIFF-FX files.